COMPARING OBJECTIVE VISUAL QUALITY IMPAIRMENT DETECTION IN 2D AND 3D VIDEO SEQUENCES

Nicolas Staelens[a] *Arnaud Boussaer*[b] *Nick Vercammen*[a]
Geert Van hoogenbemt[b] *Brecht Vermeulen*[a] *Piet Demeester*[a]

[a]Ghent University - IBBT, Department of Information Technology, Ghent, Belgium
[b]Ghent University College, Department INWE, Ghent, Belgium

ABSTRACT

Thanks to the availability of 3D-capable televisions and blu-ray players, 3D content is made accessible in the home. Recently, an extension of the H.264/AVC video coding standard has been defined for encoding 3D video content. This extension, called Multiview Video Coding, allows inter-view prediction resulting in a better compression efficiency. However, due to these inter-view dependencies impairments in one view caused by e.g. packet losses can lead to degradations in other views. Research has already been conducted towards estimating packet loss visibility in H.264/AVC encoded sequences. In this paper, we investigate the possibility of using an existing decision tree-based classifier for estimating impairment visibility in 3D MVC encoded sequences. Our results show that, in the case of losing entire pictures, it is possible to estimate packet loss visibility in 3D MVC encoded sequences with a high accuracy by only taking into account a limited number of parameters.

Index Terms— 3D, MVC, QoE, Subjective Quality, Impairment Visibility

1. INTRODUCTION

Over the past several years, 3D video has become much more popular resulting from the release of an increasing number of blockbuster movies. Watching a 3D movie positively influences end-users' overall viewing experience and makes it easier for the viewers to immerse themselves in the storyline [1]. Furthermore, watching 3D content is no longer exclusively attributed to movie theatres due to the broader availability of consumer-grade 3D television sets and blu-ray players, and the introduction of new enhanced video services such as 3D Video-on-Demand (VoD).

For storage and transmission of 3D and free viewpoint video, the Multiview Video Coding (MVC) [2] extension of the H.264/AVC standard has recently been defined. Amongst other, MVC relies on inter-view prediction in order to improve compression efficiency. As such, in the case of stereoscopic video, the right view is predicted based on the base (left) view as shown in Figure 1. Additionally, the base view in MVC is also backwards compatible with H.264/AVC.

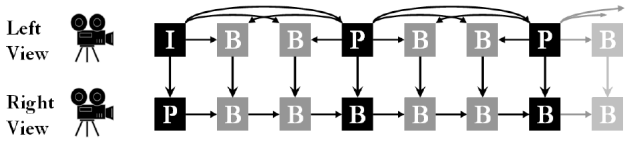

Fig. 1. Based on inter-view prediction in MVC, the right view is predicted based on the left view [2].

While streaming video over Internet Protocol (IP) -based networks, impairments (such as packet loss) can severely degrade the perceived quality of the end-users [3]. The latter is also commonly referred to as Quality of Experience (QoE) [4]. In order to maintain customer satisfaction and adequate QoE at all times, service providers should continuously monitor the quality of their video streams. This can be achieved by means of objective video quality metrics [5].

In the case of 2D video, research has already been conducted to determine whether packet losses will result in visible impairments [6, 7, 8]. However, due to the inter-view dependencies employed in MVC, packet losses in the base (left) view will also result in distortions in the right view [9]. Hence, existing models for predicting packet loss visibility in 2D video might need some additional fine-tuning before being applied to estimate impairment visibility in 3D video sequences.

In this paper, we investigate the difference in impairment visibility between 2D and 3D video sequences. Based on data collected from a new subjective experiment using 3D videos, we first try to predict packet loss visibility using an already existing no-reference bitstream-based 2D visual quality impairment detector [10]. For modeling packet loss visibility, we only consider parameters which can easily be extracted from the encoded bitstream. Next, we analyze whether the existing decision tree-based classifier should be fine-tuned in order to increase the prediction accuracy of visual impairments in 3D video sequences.

The remainder of this paper is structured as follows. In

Section 2, we start by providing an overview of already conducted subjective experiments to assess the influence of packet loss and error concealment strategies on the perceived quality of 3D videos. Next, in Section 3, we briefly discuss how packet loss visibility has been modeled so far in the case of 2D video. Section 4 describes the subjective experiment we conducted to obtain our ground-truth data for modeling packet loss visibility in 3D encoded sequences. The results of this experiment are presented in Section 5. This data is then used to investigate the possibility of using an existing decision tree-based classifier to estimate impairment visibility in 3D MVC encoded video sequences. Finally, we conclude the paper and present future work.

2. INFLUENCE OF PACKET LOSS ON PERCEIVED QUALITY IN 3D VIDEOS

Recently, a number of subjective studies have already been conducted to assess the influence of packet loss on perceived visual quality in 3D video sequences.

In [11], a subjective study was conducted to determine the most appropriate error concealment strategy to be used in the case of transmission errors in 3D sequences. Results indicate that switching to a 2D representation when errors only occur in one view is preferred in order to maintain the highest perceived visual quality. A similar study was performed in [12] which also showed that users tend to prefer switching from 3D to 2D in the case of entire frame losses. This is especially true for longer error bursts. Furthermore, it has been shown [9] that disparity negatively influences perceived 3D quality in the event of missing frames in the auxiliary view.

In [13], Guttiérrez *et al.* compare the impact of transmission errors in monoscopic and side-by-side 3D audiovisual sequences. Based on a statistical analysis, it is shown that 3D content is rated slightly better quality compared to the 2D video, except in the case of lost video packets. However, the authors state that this is mainly due to the fact that lost video packets in side-by-side 3D videos cause blockiness artifacts in different regions of each view. This, in turn, could hinder the fusion of both views leading to visual discomfort. Preference towards 3D or 2D content is also influenced by the encoding bitrate [14].

3. MODELING PACKET LOSS VISIBILITY IN 2D VIDEO SEQUENCES

In this paper, we are interested in modeling packet loss visibility by means of a binary classification in the case of 3D video sequences. Similar research has already been conducted for 2D content.

Two different approaches have been used to model impairment visibility. Kanumuri *et al.* [6] and Argyropoulos *et al.* [8] both use *regression analysis* to predict the probability that packet losses will result in impairments deemed visible to the average end-user. In the former, this probability is estimated based on a generalized linear model whereas in the latter, support vector regression is used to continuously predict packet loss visibility.

In a second approach, a *decision tree* is used by Reibman *et al.* [15] and Kanumuri *et al.* [16] to classify packet loss as visible or invisible. In this case, a binary classification is used instead of estimating the probability that packet loss will be visible or not. A decision tree can be regarded as a white box showing the complete internal structure of the classification process. Hence, an in-depth analysis of the tree can be performed leading to more insights. This analysis can, for example, provide more information on which parts of the bitstream should be protected more against possible losses in the network.

In previous research, we also used a decision tree classifier to estimate packet loss visibility for High Definition H.264/AVC encoded video sequences [10]. Our results show that impairment visibility can be estimated with a high accuracy, only taking into account high level parameters which can easily be extracted solely from the received video bitstream.

Now, we want to investigate whether these findings still hold in the case of packet losses occurring in stereo MVC encoded video sequences.

4. ASSESSING PACKET LOSS VISIBILITY IN 3D VIDEOS

In order to obtain ground-truth data for modeling impairment visibility, we conducted a new subjective experiment. Different 3D video sequences were selected, encoded and impaired using several impairment scenarios. Next, these sequences were visually presented to a number of test subjects which had to evaluate the perceived quality of each sequence. In the next sections, more information is provided on the impairment generation and the subjective quality assessment methodology used during this experiment.

4.1. Source content and encoding

Seven different stereoscopic video sequences of HD resolution (1920x1080) were selected. Each sequence had a duration of 10 seconds and a frame rate of 24 Hz. The sequences were selected to span a range of different content types ranging from low motion & low spatial details to high motion & high spatial details. More details on the selected sequences are presented in Table 1.

In order to get a first impression of the difference in impairment visibility between 2D and 3D video sequences when losing entire pictures, a subset of the encoding settings used in our previous research [10] was selected as a starting point. Each sequence was encoded in Stereo High Profile using the JM Reference Software version 17.2 into an MVC compliant bitstream. A closed GOP structure with a size of 16 and two

Table 1. Description of the stereoscopic test sequences.

Sequence	Description
SRC01	Animated spinning cars.
SRC02	Sheep eating grass.
SRC03	Overview of London.
SRC04	Racing cars.
SRC05	Car drifting in gravel.
SRC06	Close-up of two characters talking
SRC07	Close-up of two people opening bottle.

B-pictures was used. Each picture was encoded using a single slice. As we are interested in assessing the influence of network impairments, the encoding bitrate was set high enough to ensure no encoding artifacts were present in the sequences. On average, the encoding bitrate was equal to 12 Mbps. After encoding, each sequence was also visually inspected.

4.2. Impairment generation

Impairments in the MVC encoded sequences were injected by dropping particular slices (= pictures) using our in-house developed open-source multimedia streamer 'Sirannon' [17].

First, single frame losses were considered by dropping an I-, P- or B-picture in each view separately. This results in 5 different single loss impairment scenarios[1]. Next, multiple picture losses were also considered by dropping up to four consecutive slices (in encoding order). An overview of these different impairment scenarios is provided in Figure 2. As such, these impairments influence both views simultaneously.

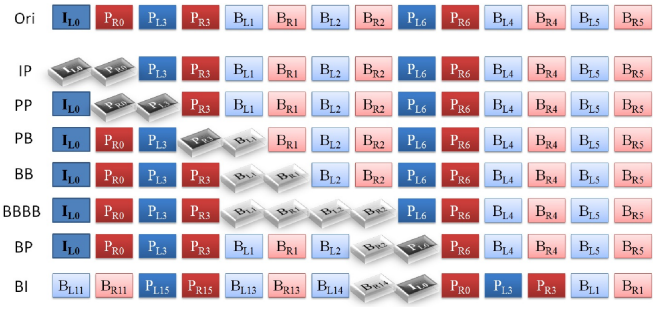

Fig. 2. Overview of the seven impairment scenarios created by dropping multiple consecutive pictures.

In total, this results in 12 impairment scenarios per sequence. Including the original, 91 video sequences needed to be evaluated by our test subjects.

As concealment strategy, frame copy is used as this was also the case in our previous research. As such, the dropped picture was replaced by the last correctly received picture in the same view.

[1]I-pictures are only present in the left view.

4.3. Subjective test methodology

The sequences were presented to the test subjects using the Single Stimulus (SS) Absolute Category Rating (ACR) assessment methodology as specified in [18]. This implies that all test sequences were shown one-after-another as depicted in Figure 3.

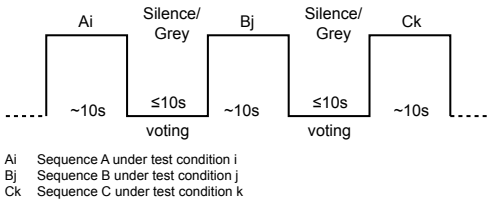

Fig. 3. Typical trail structure for an SS methodology [18], during which sequences are presented one at a time and immediately evaluated after watching.

Before the start of the experiment, all test subjects received specific instructions on how to evaluate each video sequence. A Snellen chart and Ishihara plates were used to check each subject for normal vision. Subjects were also screened for depth perception by means of a Randot Stereo test. In order to get the subjects familiarized with the trail structure and subjective test software, three training sequences were used. These training sequences also represented the type of impairments subjects could expect during the experiment. The quality ratings given to these sequences are not taken into account when processing the results.

After watching each sequence, the test subjects were required to indicate whether they perceived any kind of visual impairment by means of a simple yes/no answer.

The experiment was conducted inside an ITU-R BT.500-12 compliant test environment. A 40" Full HD LED television set was used to display the sequences. Test subjects were seated at a distance of three times the display height (3H) and were given the active shutter glasses which came with the 3D television. The screen was also calibrated to counteract the reduced luminance caused by the use of these active shutter glasses.

4.4. Participants

A total number of 24 non-expert subjects participated in this experiment. Based on the post-experiment screening procedure outlined in Annex V of the VQEG HDTV report [19], we ensured no outliers were present in our subjective data.

Amongst all participants, 17 were male test subjects with ages ranging from 18 to 57 years old (average age: 30). The average age of the female subjects was 31 years old with a minimum and maximum age of respectively 18 and 50.

All subjects evaluated the 91 video sequences. Randomisation was used so that no two subjects evaluated the sequences

in exactly the same order. The experiment took around half an hour to complete.

5. ESTIMATING IMPAIRMENT VISIBILITY IN 3D MVC ENCODED VIDEO SEQUENCES

In this section, we present the results of our subjective experiment. First, we try to estimate impairment visibility in 3D stereoscopic videos based on an already existing classifier and analyze whether fine-tuning is needed in order to increase performance and accuracy.

In our previous research [10], the high level parameters listed in Table 2 were extracted from H.264/AVC encoded bitstreams and used to construct a no-reference bitstream-based decision tree for classifying packet loss visibility as depicted in Figure 4.

Table 2. Overview of the high level parameters extracted from H.264/AVC encoded bitstreams for modeling impairment visibility.

Parameter	Description
imp_pic_type	Type of picture (I, P or B) where the loss originates from.
imp_in_gop_pos	Position of the impaired picture within the corresponding GOP.
imp_cons_slice_drops	Number of consecutive slice drops.
imp_cons_b_slice_drops	Number of consecutive B-slice drops.

This classifier was trained and validated based on data obtained from a subjective experiment. Packet losses were labeled as visible in case 75% or more of the test subjects noticed the impairment. Starting from a detection threshold of 75%, Mean Opinion Scores (MOS) also drop below 4 [10]. Results showed that impairment visibility can be estimated with a high accuracy, only taking into account these high level parameters. Furthermore, this classifier can also be applied to sequences encoded with multiple slices per picture.

Applying the classifier depicted in Figure 4 to our newly obtained subjective data results in a classification accuracy of 85.7%. As such, our existing classifier trained using H.264/AVC encoded sequences can be used to estimate packet loss visibility with a high accuracy for our 3D MVC encoded videos.

The performance of a decision tree can also be analyzed in terms of the True Positive (TP) rates. The TP rate for visible impairments indicates the percentage of packet losses correctly classified as visible to the average end-user. Likewise, the TP rate for invisible losses refers to the percentage of packet losses correctly classified as invisible. A comparison between the accuracy and TP rates for classifying packet loss visibility in our H.264/AVC and MVC encoded sequences is presented in Table 3. It can be seen that the TP rate for visible impairments is significantly higher for our MVC encoded sequences. This means that the classifier is able to predict the occurrence of visible impairments with a high accuracy. The TP rate for invisible impairments is only slightly lower compared to the H.264/AVC encoded sequences. In general, high TP rates for visible impairments are preferred as in the case of real-time monitoring it is important not to misclassify packet losses as being invisible to often. It must be noted that the H.264/AVC encoded sequences included much more encoder configuration settings [10] (e.g. multiple B-pictures, multiple slices per picture, ...).

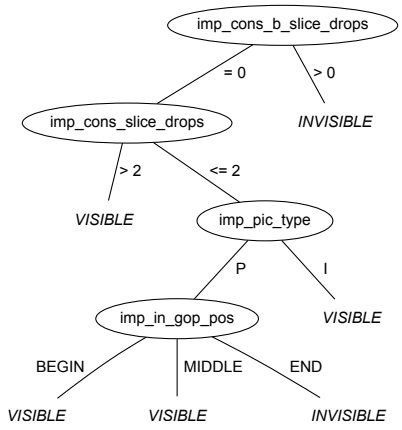

Fig. 4. Decision tree for classifying the visibility of packet losses in HD H.264/AVC encoded video sequences.

Table 3. Performance comparison between estimating impairment visibility in H.264/AVC and MVC encoded video sequences.

	H.264/AVC sequences	MVC encoded sequences
Overall accuracy	83.1%	85.7%
TP rate visible	84.0%	92.1%
TP rate invisible	82.1%	80.4%

Our impairment scenarios included losses in one view at a time and in the two views simultaneously. As can be seen in Figure 2, losses in both views were inserted by dropping consecutive pictures in encoding order. This can result in dropping two pictures of different types (e.g. a P-picture followed by a B-picture). In that case, the parameter `imp_pic_type` is set equal to the most dominant picture type. For example, if a P- and B-picture are dropped, `imp_pic_type` is set to be a P-picture. Estimating packet loss visibility in sequences with losses only occurring in both views simultaneously results in an overall classification accuracy of 89.8% with TP rates for visible and invisible impairments 88.0% and 91.7%, respectively.

Analyzing our subjective data in more details shows that losses of I-pictures are visible in 100% of the cases. Losses

of one or two consecutive B-pictures are never perceived by our test subjects, even if the losses occur simultaneously in both views. This perfectly corresponds with the classification process of our decision tree shown in Figure 4. As such, only some classification errors occur in the case of losing P-pictures. Training a decision tree using only the obtained data corresponding with losses occurring in P-pictures results in the simple tree depicted in Figure 5. Again, this matches with the subtree of the branch '`imp_pic_type = P`' from Figure 4.

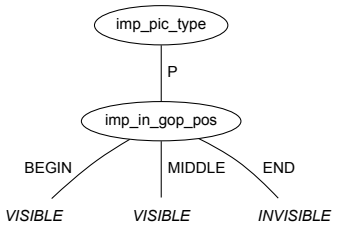

Fig. 5. Decision tree for classifying packet loss visibility in the case of losing P-pictures in 3D MVC encoded sequences.

Therefore, based on these findings, we can conclude that the decision tree depicted in Figure 4 (trained using H.264/AVC encoded videos) is the optimal tree for classifying packet loss visibility when losing entire pictures in our 3D MVC encoded video sequences.

6. CONCLUSION AND FUTURE WORK

In this paper, we investigated the difference in impairment visibility between 2D H.264/AVC and 3D MVC encoded video sequences. Several stereoscopic video sequences were encoded and impaired by dropping particular pictures in the encoded bitstream. These impaired sequences were, in turn, used to conduct a subjective experiment in order to gather ground-truth data concerning packet loss visibility in 3D content. In previous research, we already constructed a no-reference bitstream-based visual quality impairment detector to determine packet loss visibility in HD H.264/AVC encoded sequences. Based on the data obtained from the subjective experiment conducted as part of the research presented in this paper, we investigated the possibility of using this existing classifier to predict impairment visibility in 3D MVC encoded videos.

Our results show that it is possible to predict packet loss visibility in MVC encoded video sequences using a decision tree-based classifier trained on HD H.264/AVC content. A high prediction accuracy is obtained by only taking into account a limited number of parameters which can easily be extracted from the encoded bitstream without the need for decoding or deep packet inspection.

In this research, only a subset of the encoding settings was considered for encoding the different stereoscopic source videos. In particular, we only considered the influence of dropping entire pictures. Additional research is needed to create a larger pool of test sequences using multiple encoding settings. Increasing the number of slices per pictures and using a different amount of B-pictures between two reference pictures might be interesting to consider in future work. This way, a dataset can be obtained similar to the HD H.264/AVC-based dataset used to construct our proposed decision tree classifier. Last, different error concealment techniques can also be taken into account.

7. ACKNOWLEDGMENT

Part of the research activities described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT) and the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT). This paper is the result of research carried out as part of the OMUS project funded by the IBBT. OMUS is being carried out by a consortium of the industrial partners: Excentis, Streamovations, Technicolor and Televic in cooperation with the IBBT research groups: IBCN, WiCa & Multimedia Lab (UGent), SMIT (VUB), PATS (UA) and COSIC (KUL).

8. REFERENCES

[1] N. Staelens, K. Casier, W. Van den Broeck, B. Vermeulen, and P. Demeester, "Determining customer's willingness to pay during in-lab and real-life video quality evaluation," *Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-12)*, January 2012.

[2] A. Vetro, T. Wiegand, and G.J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard," *Proceedings of the IEEE*, vol. 99, no. 4, pp. 626 –642, April 2011.

[3] M.H. Pinson, S. Wolf, and G. Cermak, "HDTV subjective quality of H.264 vs. MPEG-2, with and without packet loss," *IEEE Transactions on Broadcasting*, vol. 56, no. 1, pp. 86–91, March 2010.

[4] ITU-T Recommendation P.10/G.100 Amd 2, "Vocabulary for performance and quality of service," 2008.

[5] N. Staelens, I. Sedano, M. Barkowsky, L. Janowski, K. Brunnström, and P. Le Callet, "Standardized toolchain and model development for video quality assessment - the mission of the joint effort group in VQEG," in *Third International Workshop on Quality of Multimedia Experience (QoMEX)*, September 2011.

[6] S. Kanumuri, S.G. Subramanian, P.C. Cosman, and A.R. Reibman, "Predicting H.264 packet loss visibility using a generalized linear model," in *International Conference on Image Processing*, October 2006, pp. 2245–2248.

[7] N. Staelens, N. Vercammen, Y. Dhondt, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, "ViQID: A No-Reference Bit Stream-based Visual Quality Impairment Detector," in *Second International Workshop on Quality of Multimedia Experience (QoMEX)*, June 2010, pp. 206 –211.

[8] S. Argyropoulos, A. Raake, M. Garcia, and P. List, "No-reference video quality assessment for SD and HD H.264/AVC sequences based on continuous estimates of packet loss visibility," in *Third International Workshop on Quality of Multimedia Experience (QoMEX)*, September 2011, pp. 31–36.

[9] L. Pinto, J. Carreira, S. Faria, N. Rodrigues, and P. Assuncao, "Subjective quality factors in packet 3D video," in *Third International Workshop on Quality of Multimedia Experience (QoMEX)*, September 2011, pp. 149 – 154.

[10] N. Staelens, G. Van Wallendael, K. Crombecq, N. Vercammen, J. De Cock, B. Vermeulen, R. Van de Walle, T. Dhaene, and P. Demeester, "No-Reference Bitstream-based Visual Quality Impairment Detection for High Definition H.264/AVC Encoded Video Sequences," *IEEE Transactions on Broadcasting*, vol. 58, no. 2, pp. 187–199, June 2012.

[11] M. Barkowsky, K. Wang, R. Cousseau, K. Brunnström, R. Olsson, and P. Le Callet, "Subjective quality assessment of error concealment strategies for 3DTV in the presence of asymmetric transmission errors," in *18th International Packet Video Workshop (PV)*, December 2010, pp. 193 –200.

[12] J. Carreira, L. Pinto, N. Rodrigues, S. Faria, and P. Assuncao, "Subjective assessment of frame loss concealment methods in 3D video," in *Picture Coding Symposium (PCS)*, December 2010, pp. 182 –185.

[13] J. Gutiérrez, P. Pérez, F. Jaureguizar, J. Cabrera, and N. García, "Subjective evaluation of transmission errors in IPTV and 3DTV," in *IEEE Visual Communications and Image Processing (VCIP)*, November 2011, pp. 1 –4.

[14] K. Brunnström, I. Sedano, K. Wang, M. Barkowsky, M. Kihl, B. Andrén, P. Le Callet, M. Sjöström, and A. Aurelius, "2D No-Reference Video Quality Model Development and 3D Video Transmission Quality," *Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-12)*, January 2012.

[15] A.R. Reibman, S. Kanumuri, V. Vaishampayan, and P.C. Cosman, "Visibility of individual packet losses in MPEG-2 video," in *International Conference on Image Processing*, October 2004, vol. 1, pp. 171–174.

[16] S. Kanumuri, P.C. Cosman, A.R. Reibman, and V.A. Vaishampayan, "Modeling packet-loss visibility in MPEG-2 video," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 341–355, April 2006.

[17] A. Rombaut, N. Staelens, N. Vercammen, B. Vermeulen, and P. Demeester, "xStreamer: Modular Multimedia Streaming," in *Proceedings of the seventeenth ACM international conference on Multimedia*, 2009, pp. 929–930.

[18] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union (ITU), 1999.

[19] Video Quality Experts Group (VQEG), "Report on the Validation of Video Quality Models for High Definition Video Content," June 2010.